VIROLOGY 249, 219–230 (1998)
ARTICLE NO. VY989339

Characterization of the Initiator and Downstream Promoter Elements of Herpes Simplex Virus 1 Late Genes

Amy M. Woerner and Jerry P. Weir[1]

Division of Viral Products, Center for Biologics Evaluation and Research, Food and Drug Administration, 1401 Rockville Pike, Bethesda, Maryland 20892

Received January 16, 1998; returned to author for revision March 9, 1998; accepted July 16, 1998

Previously identified *cis*-acting regulatory elements of herpes simplex virus (HSV) 1 late promoters include a TATA element upstream from the start of transcription, an initiator-like element at the start of transcription, and sequences downstream from the start of transcription. To determine whether these elements are functionally equivalent to similar elements from other eukaryotic genes, model late promoters were constructed using well-characterized regulatory elements from non-HSV genes. These modular promoters were then inserted into the viral genome upstream from a *lacZ* marker gene. Results showed that a eukaryotic initiator element, along with a TATA element, can function as a late HSV promoter. Several initiator sequences from both viral and nonviral genes were functionally similar to the initiator-like element in HSV-1 late promoters; however, a random sequence of the same size and a similarly located sequence from the HSV-1 early thymidine kinase promoter could not substitute for the initiator element. These results indicate that eukaryotic initiator elements are functionally equivalent to HSV-1 late promoter initiator elements. In addition, the downstream element of the late glycoprotein C promoter was further analyzed by construction of a series of small deletions and insertions. The presence of the downstream glycoprotein C region in a promoter consisting of a strong TATA and initiator element increased mRNA expression by a modest amount; this effect appeared to be sequence specific and dependent on its exact alignment with the upstream elements of the promoter.

INTRODUCTION

Transcription of the >75 known HSV-1 genes is coordinately regulated at the level of transcription (for a review, see Roizman and Sears, 1996). The five immediate-early (IE) genes are expressed first, and their expression occurs in the absence of *de novo* protein synthesis. The subsequent expression of early genes and late genes is dependent on production of functional IE proteins; additionally, expression of late genes requires viral DNA replication. The product of one late gene (UL48), designated as VP16, is packaged in the virion and strongly transactivates expression of the IE genes. Although the general scheme of viral gene expression has been known for some time, the mechanisms that control this well-orchestrated process are still unclear.

Detailed analysis of the promoter regions for various viral genes has elucidated at least some of the basic features of promoters from each of the temporal classes. Although IE promoters contain common eukaryotic *cis*-acting regulatory elements such as a TATA element and Sp1 response elements, the most distinguishing feature of this type of promoter is the presence of a *cis*-acting regulatory sequence, TAATGARAT, that is responsible for the transactivation effect of VP16 (for review, see O'Hare, 1993). VP16 itself does not bind DNA, but it interacts with several cellular factors, including Oct-1, to form a complex that directly binds to the TAATGARAT sequence. In addition, some IE promoters contain a binding site for the IE protein ICP4 at the start of transcription; binding of ICP4 to this site represses transcription (Michael and Roizman, 1993; Roberts *et al.,* 1988). In contrast to IE promoters, early promoters do not appear to contain virus-specific *cis*-acting regulatory elements. Studies with the thymidine kinase (tk) promoter identified a TATA element, Sp1 binding sites, and a CCAAT box, all of which are common eukaryotic transcriptional regulatory elements (Coen *et al.,* 1986). Indeed, some cellular promoters, when inserted into the viral genome, function as early viral promoters (Panning and Smiley, 1989). This observation reinforced the concept that herpes simplex virus (HSV) early promoters were typical eukaryotic promoters.

Analysis of several late promoters has revealed another type of promoter architecture, but one also lacking any defined virus-specific regulatory sequence elements (Flanagan *et al.,* 1991; Guzowski and Wagner, 1993; Homa *et al.,* 1986; Kibler *et al.,* 1991; Mavromara-Nazos and Roizman, 1989; Steffy and Weir, 1991). Unlike early promoters, there do not appear to be any *cis*-acting regulatory elements upstream from the TATA element

[1] To whom reprint requests should be addressed. Fax: (301) 480-6124.

0042-6822/98

(Homa *et al.,* 1986; Johnson and Everett, 1986). However, sequences at the start of transcription and further downstream appear to be necessary for maximal expression from these promoters (Guzowski *et al.,* 1994; Mavromara and Roizman, 1989; Weir and Narayanan, 1990). The sequence at the start of transcription strongly resembles the eukaryotic initiator element (Inr) (Steffy and Weir, 1991); the sequence element further downstream has been termed the downstream activation sequence (DAS), and several similar sequence elements have been identified in HSV-1 late genes (Guzowski *et al.,* 1994).

From these and other studies, a general model for late HSV-1 promoters has emerged (for review, see Wagner *et al.,* 1995). The late promoter consists of a core promoter from the TATA element through the start of transcription, linked to a DAS within the nontranslated leader sequence (Guzowski and Wagner, 1993). However, there are still some questions that need to be addressed regarding the regulatory elements that constitute these promoters. Although the initiator-like sequence of HSV-1 late promoters bears a strong resemblance to an eukaryotic Inr and, in fact, can direct transcription *in vitro* when linked to a TATA element (Gu and DeLuca, 1994), the functional equivalence of these elements during virus infection has not been demonstrated. To address this, model late promoters were constructed using the regulatory elements of several well-studied eukaryotic promoters. These modular promoters then were inserted into the HSV-1 genome and evaluated for their ability to direct transcription. Furthermore, although extensive mutational analysis of the late glycoprotein C (gC) promoter has also indicated the presence of a downstream activating element, the effects of mutations in this region on expression were much less pronounced than similar mutations in the DAS of the UL38 late gene (Guzowski *et al.,* 1994; Guzowski and Wagner, 1993; Steffy and Weir, 1991; Weir and Narayanan, 1990). To further understand the role and the contribution of the gC downstream element in late gene expression, a more extensive analysis of this region of the promoter was undertaken. Model promoters were constructed that contained various combinations of *cis*-acting regulatory sequences to examine their effects on the expression of mRNA.

RESULTS

Promoters derived from non-HSV genes can function as late promoters in a recombinant HSV-1

Previously identified *cis*-acting regulatory elements of late HSV-1 promoters include a TATA element in the region upstream from the start of transcription (Homa *et al.,* 1986; Homa *et al.,* 1988), an initiator-like sequence at the start of transcription (Steffy and Weir, 1991), and sequences further downstream (Guzowski *et al.,* 1994; Guzowski and Wagner, 1993; Mavromara-Nazos and Roizman, 1989; Weir and Narayanan, 1990). To determine whether these late promoter elements are functionally equivalent to similar elements in other eukaryotic genes, several model promoters were constructed using well-characterized regulatory elements from non-HSV genes. These promoters were inserted upstream of an *Escherichia coli lacZ* gene and recombined into the viral genome (Fig. 1A). Because one of the best characterized

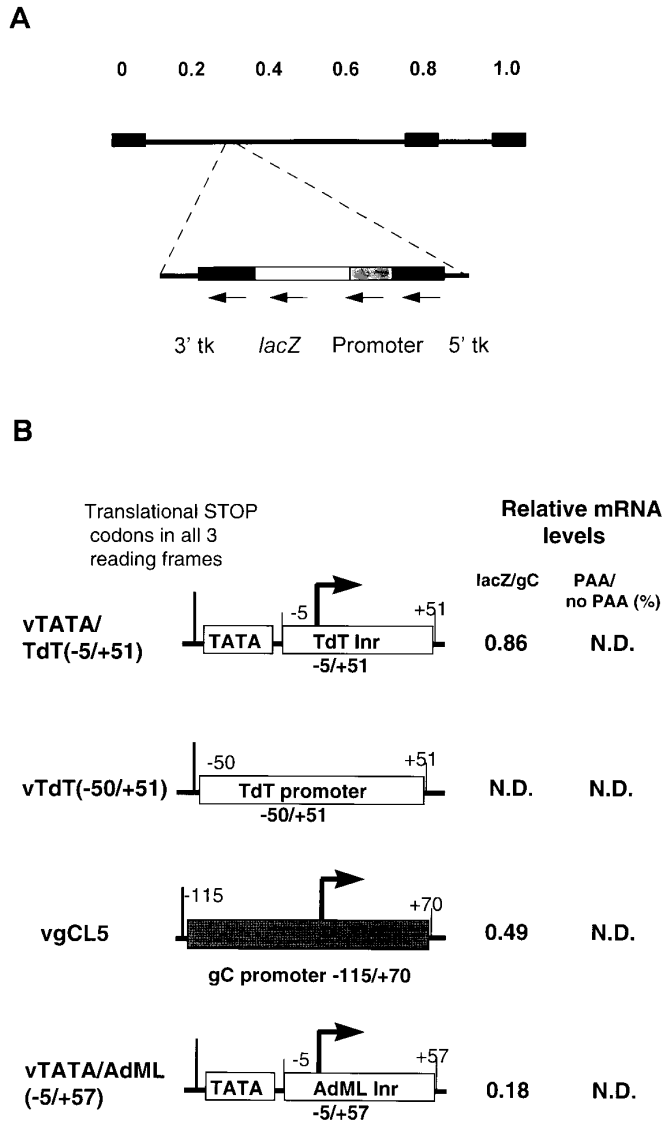


FIG. 1. Insertion of synthetic eukaryotic promoters into the HSV-1 genome. (A) Schematic of the genomic structure of recombinant viruses containing promoters linked to the *lacZ* gene. Each chimeric gene containing the described promoter and the *lacZ* gene was inserted into the tk gene by homologous recombination. (B) Promoter structure and β-gal activity of recombinant viruses that use synthetic promoters constructed from eukaryotic promoter elements to express the *lacZ* gene. Vero cells were infected with the indicated viruses at an m.o.i. of 5 in the presence or absence of PAA. RNA was isolated at 12 h p.i. and analyzed by primer extension for gC and *lacZ* mRNAs. The gC and *lacZ* primer extension products were quantified with a PhosphorImager (Molecular Dynamics, Sunnyvale, CA), and the level of *lacZ* mRNA was expressed relative to that of gC. (N.D.) Primer extension product not detected.

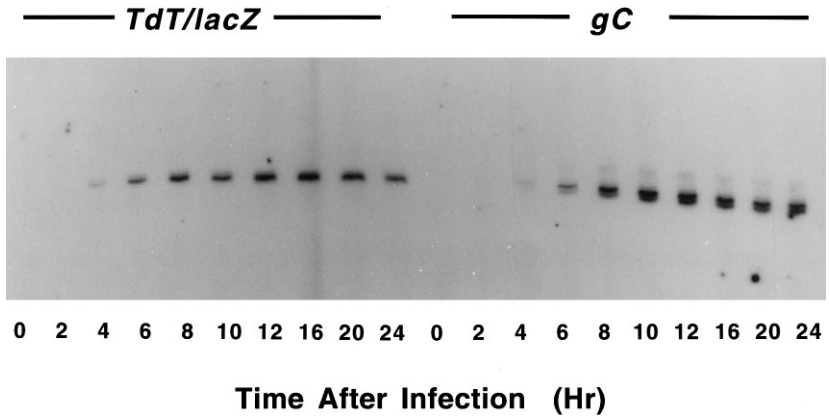


FIG. 2. Time course of *lacZ* and gC expression in vTATA/TdT(−5/+51)-infected cells. Vero cells were infected with vTATA/TdT(−5/+51) at an m.o.i. of 5, and RNA was isolated at the indicated times. RNA was hybridized to either a gC- or a *lacZ*-specific primer, extended with reverse transcriptase, and run on an 8% sequencing gel. The expected primer extension product from the TATA/TdT promoter is 74 bases; the expected gC primer extension product is 70 bases.

eukaryotic initiators is that of the murine terminal deoxynucleotidyl transferase (TdT) gene, we linked the TdT initiator and downstream sequences from −5 to +51 [the translation initiation codon of the natural TdT gene is at +52 (Smale and Baltimore, 1989)] to a strong TATA element from the adenovirus major late (AdML) promoter and inserted this synthetic promoter upstream of the *lacZ* gene. A recombinant virus containing this promoter expressed *lacZ* mRNA, as determined by primer extension analysis. The *lacZ* mRNA expression, like that of the late gC mRNA, was dependent on viral DNA replication as shown by inhibition in the presence of phosphonoacetic acid (PAA) [Fig. 1B, vTATA/TdT(−5/+51)]. By comparison, in the presence of PAA, the early tk mRNA increased ~2.5-fold from 6 to 12 h p.i. (data not shown). The level of *lacZ* mRNA expressed in Vero cells 12 h after infection with vTATA/TdT(−5/+51) was ~18-fold higher than that expressed on infection of cells with vgCL5 (Weir and Narayanan, 1990), a recombinant virus that uses the HSV-1 gC promoter from −115/+70 to direct *lacZ* expression. A second modular promoter was constructed that replaced the TdT sequences with the AdML initiator and downstream sequences from −5 to +57. This recombinant virus also expressed *lacZ* mRNA that was dependent on viral DNA replication [Fig. 1B, vTATA/AdML(−5/+57)], although the level of *lacZ* mRNA expressed was lower than that of either vTATA/TdT(−5/+51) or vgCL5. In contrast to the expression of *lacZ* mRNA by these modular promoters in recombinant viruses, the TATA-less TdT promoter from −50 to +51 did not direct detectable expression of the *lacZ* gene [Fig. 1B, v/TdT(−50/+51)]. Taken together, these data indicate that an eukaryotic Inr element, along with a TATA element, function as a late promoter in the HSV-1 genome. In contrast, the strong eukaryotic Inr from the TdT promoter, in its natural TATA-less promoter context, is ineffective in the viral genome.

Although *lacZ* mRNA directed from these modular promoters was dependent on viral DNA replication as shown by PAA sensitivity, the late promoter activity of the TATA/TdT synthetic promoter was also investigated by a temporal analysis of mRNA expression. Vero cells were infected with the recombinant virus vTATA/TdT(−5/+51) and RNA isolated at various times p.i. Primer extension analysis, using primers complementary to the 5′ end of either the *lacZ* gene or the gC gene, indicated that *lacZ* mRNA expression from the TATA/TdT promoter mirrored gC mRNA expression from 2 to 24 h p.i. (Fig. 2). Thus, the temporal expression of *lacZ* mRNA from the TATA/TdT promoter also indicated that this modular promoter functioned as a late promoter in the viral genome.

Characterization of initiator elements derived from various genes

To determine whether other identified eukaryotic initiator elements were functionally similar to the initiator-like element in HSV-1 late promoters, a second set of synthetic modular promoters was constructed, inserted upstream of the *lacZ* gene, and recombined into the viral genome (Fig. 3). Each promoter in this set had the AdML TATA element, as in vTATA/TdT(−5/+51), linked to a different 16-bp sequence at the site of transcription initiation (−5/+11). Five of these sequences were derived from previously defined eukaryotic initiators, one was from the initiator-like region of the HSV-1 gC gene, one from the similarly located region of the HSV-1 tk gene, and one sequence was a random sequence previously shown to have no Inr activity in *in vitro* transcription assays (Javahery *et al.,* 1994). Because the influences of the sequences further downstream on late promoter activity are not yet completely understood, the TATA/Inr sequences were linked to the downstream region of the HSV-1 gC gene from +20 to +70. A *Sal*I restriction site (+12/+17) separated the TATA/Inr from the gC nontrans-

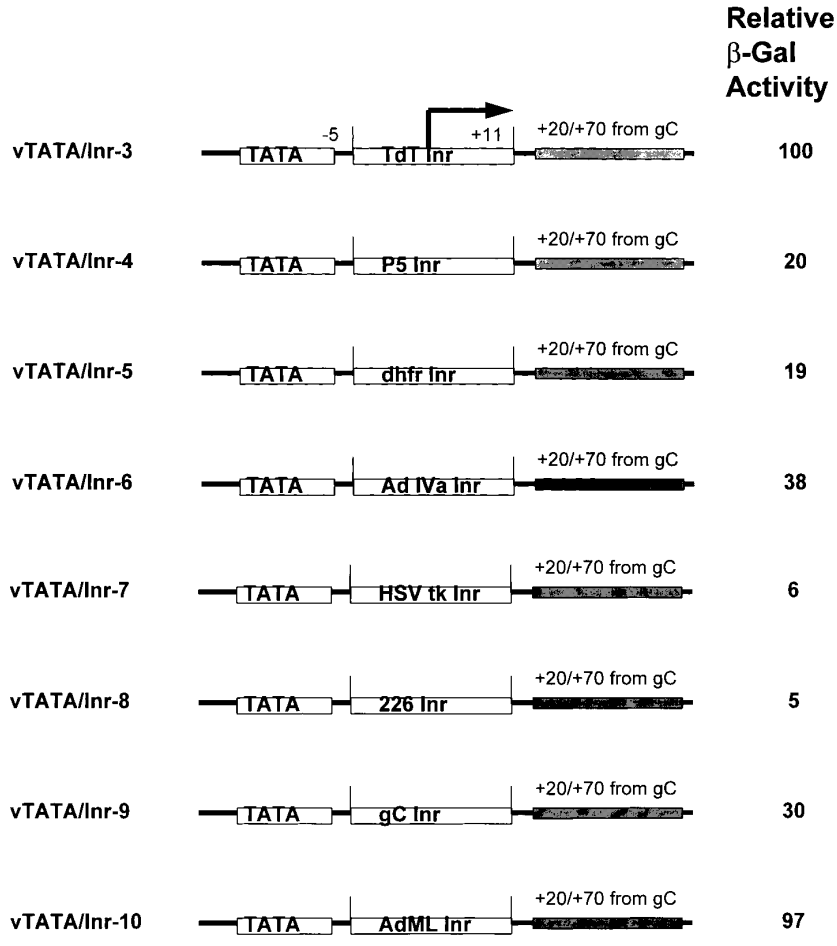


FIG. 3. Insertion of initiator sequences into a synthetic late promoter. Synthetic promoters were constructed that contained the 16-bp initiator sequence from the indicated promoters downstream from the AdML TATA element. These chimeric TATA/Inr regions were inserted upstream of the gC 5′ nontranslated region from +20 to +70, which was linked to the *lacZ* gene. Vero cells were infected with the indicated viruses at an m.o.i. of 5, and the β-gal activity was determined at 24 h p.i. and expressed relative to that of vTATA/Inr-3.

lated region such that the downstream gC sequence was only 2 bp closer to the initiator sequences in the modular promoters than in the authentic gC gene.

Each modular promoter containing a 16-bp sequence derived from a previously identified eukaryotic initiator expressed appreciable β-galactosidase (β-gal) activity. The TdT Inr and the AdML Inr sequences produced the highest levels of β-gal activity (Fig. 3, vTATA/Inr-3 and vTATA/Inr-10); the adenovirus IVa Inr element produced an intermediate level of β-gal activity (Fig. 3, vTATA/Inr-6), and the Inr elements from the adeno-associated virus P5 and the dihydrofolate reductase promoters were the weakest of the known Inr elements (Fig. 3, vTATA/Inr-4 and vTATA/Inr−5). Insertion of the initiator-like sequence from the HSV-1 gC promoter (vTATA/Inr-9) into this promoter context resulted in a promoter that produced ~30% as much β-gal activity as the TdT and AdML initiator sequences. In contrast, a random 16-bp sequence designated as 226 (Javahery *et al.,* 1994) and a similarly positioned 16-bp sequence from the start of the tk gene produced very low levels of β-gal activity when inserted into this synthetic promoter context (vTATA/Inr-8 and vTATA/Inr-7). No appreciable β-gal activity was detected from any promoter when virus infections were done in the presence of PAA (data not shown).

Although the relative β-gal activity most likely reflects the relative transcriptional activity from each promoter construct due to the fact that each gene is identical downstream from +11, analysis of *lacZ* mRNA produced from recombinant virus infection provided a more direct measure of transcriptional activity. RNA was isolated from vTATA/Inr −3, −7, −8, −9, and −10 infected cells (TdT, HSV-1 tk, 226, HSV-1 gC, and AdML initiators, respectively), and analyzed by primer extension using primers complementary to the 5′ regions of *lacZ* and gC genes (Fig. 4). The results from this analysis showed that the relative levels of *lacZ* mRNA from the synthetic promoters roughly reflected the β-gal activity measured. In addition, RNA was isolated from vTATA/Inr-3- and vTATA/TdT(−5/+51)-infected cells and analyzed by primer extension using a sequence ladder generated from the corresponding plasmid DNA and the same primer. The

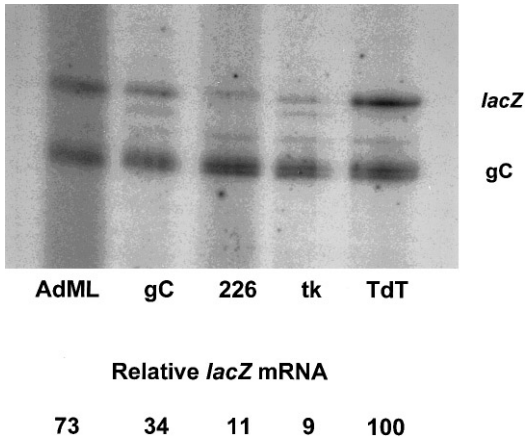


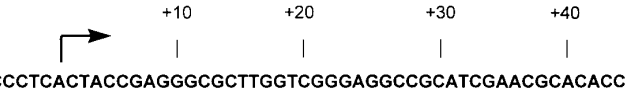
FIG. 4. Primer extension analysis of infected cell mRNA. RNA was isolated from Vero cells infected with vTATA/Inr-10 (AdML Inr), vTATA/Inr-9 (HSV-1 gC Inr), vTATA/Inr-8 (226 sequence), vTATA/Inr-7 (HSV-1 tk sequence), or vTATA/Inr-3 (TdT Inr); hybridized to both gC- and *lacZ*-specific primers in the same reaction; and extended with reverse transcriptase. The gC and *lacZ* primer extension products were quantified with a Molecular Dynamics PhosphorImager. The levels of the *lacZ* mRNAs were expressed relative to that of vTATA/Inr-3, after normalization to the level of gC in each reaction.

start of *lacZ* mRNA in both viruses mapped to the same site and was identical to the authentic transcription initiation site of the TdT promoter (data not shown). Taken together, these results indicate that other eukaryotic initiator elements can functionally replace the similarly located sequence in an HSV-1 late promoter and support the concept that an eukaryotic initiator element is an integral part of the HSV-1 late promoter.

Analysis of the downstream region of the gC promoter

Previous analysis of the gC promoter using 3′ deletions and a series of point mutations revealed that the sequences downstream from the start of transcription contributed to the magnitude of expression from this promoter but did not affect temporal expression (Steffy and Weir, 1991; Weir and Narayanan, 1990). In one analysis, two separate 4-bp mutations between +31 and +39 lowered gene expression by 33–41%. Although this was a modest reduction in activity, it was in the same region implicated as necessary for maximal activity by 3′ deletion analysis. To further characterize the role of this sequence in gene expression, a series of recombinant viruses containing small deletions and insertions in the downstream region of the gC promoter were constructed, and the relative levels of *lacZ* mRNA were compared (Figs. 5 and 6). These mutants were constructed by using the unique restriction sites in the downstream gC promoter region that were generated during a linker-scanning mutational analysis of the gC promoter in gCL5 (Fig. 1B); several of these sites were found to have a negligible effect on *lacZ* gene expression (Steffy and Weir, 1991). Using these unique restriction sites, a 15-bp deletion from +29 to +43 (v21/26) was constructed in the gC promoter, and the relative level of *lacZ* mRNA produced from v21/26-infected cells was compared with that in vgCL5-infected cells (Fig. 6A). This 15-bp mutation reduced *lacZ* mRNA by 43%, whereas a second 15-bp deletion between +13 and +27 (v23/21) had virtually no effect, suggesting that effect of the +29/+43 deletion was sequence specific. Further evidence for the sequence-specific effect of this region on expression was obtained by the insertion of a random 13-bp sequence into v21/26 at the site of deletion (v21/26i4). The addition of a random 13 bp in the same region did not restore *lacZ* mRNA to the unmutated promoter level, indicating that although the effect of the 15-bp region on expression from the gC promoter may be modest, it appears to be sequence specific.

In contrast to the deletions in v21/26 and v23/21, a 10-bp deletion between +10 and +19 (v31/27) resulted in a >3-fold increase in *lacZ* mRNA from the major start site of the gC promoter, as well as an increase in the mRNA initiating at the minor gC start sites (Fig. 6A). Both the major and minor *lacZ* mRNAs from v31/27



VIRUS	MUTATION	RELATIVE mRNA LEVEL (%)
21/26	Δ15 bp (+29/43)	57±2
21/26i4	Δ15 bp (+29/+43)/ insertion of 13 bp	67±15
23/21	Δ15 bp (+13/+27)	108±15
31/27	Δ10 bp (+10/+19)	315±36
31/27-1	Δ10 bp (+10/+19)/ insertion of 10 bp	193±29
27/21	Δ9 bp (+20/+28)	223±61
27+5	5 bp insertion at +20	58±16
27+10	10 bp insertion at +20	74±15
27+23	23 bp insertion at +20	166±41

FIG. 5. Summary of deletions and insertions in the downstream gC region. Recombinant viruses were constructed from previously described linker-scanning mutants of vgCL5 (gC sequences from −115 to +70); the deleted base-pairs for each recombinant are indicated. Vero cells were infected with each recombinant virus, and RNA was isolated 24 h p.i., analyzed by primer extension assay, and quantified with a Molecular Dynamics PhosphorImager. For viruses with deletions, the levels of the *lacZ* mRNAs were expressed relative to that of vgCL5, after normalization to the level of gC in each reaction. For the viruses with insertions at +20, the level of the *lacZ* mRNAs was expressed relative to that of vgCP27.

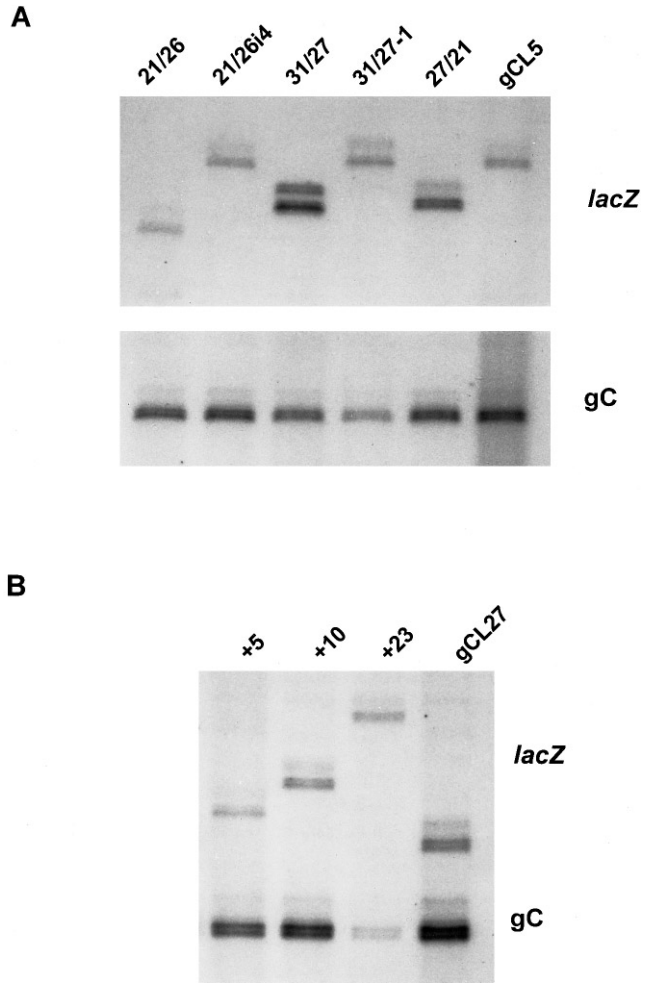


FIG. 6. Primer extension analysis of infected cell mRNA. (A) Representative primer extension gel showing *lacZ* and gC mRNA extension products from recombinant viruses containing deletions in the 5′-noncoding region of the gC gene. (B) Representative primer extension gel showing *lacZ* and gC mRNA extension products from recombinant viruses containing 5-, 10-, or 23-bp insertions at +20 in the 5′-noncoding region of the gC gene.

originated from the same sites as gC mRNA as determined by running the primer extension products alongside a sequencing ladder (data not shown). To determine whether the +10/+19 deletion removed a negative regulatory sequence, a random 10-bp sequence was inserted at the site of deletion (v31/27–1). This insertion resulted in a decrease in the relative level of *lacZ* mRNA from that of v31/27, but the level of *lacZ* mRNA was still 2-fold greater than in the unmutated promoter (Figs. 5 and 6). Because 10 bp is approximately one turn of the DNA helix, this suggests that the removal of 10 bp between +10 and +19 may have had a positional effect on a positive downstream element as well as eliminated a potentially negative regulatory element. A 9-bp deletion between +20 and +28 (v27/21) also increased the relative level of *lacZ* mRNA by 2-fold, providing further support for the idea of a positional effect of a positive-acting downstream element.

To further explore the possibility that positioning of a downstream sequence element is important for its effect on the gC promoter activity, random 5-, 10-, and 23-bp sequences were inserted into the unique *Sal*I site of vgCL27 (+14/+19) (Fig. 5). The relative levels of *lacZ* mRNA were determined by primer extension analysis and compared with that of vgCL27 (Fig. 6B). Insertion of sequences corresponding to one half or one turn of the DNA helix lowered *lacZ* mRNA expression by ~42% and ~26% (Fig. 5B, lanes +5 and +10), whereas insertion of sequences corresponding to two turns actually increased mRNA levels by 67% (Fig. 5B, lane +23). Taken together, these results suggest that the position of the downstream sequence element is important for its effect on gC promoter activity.

The effect of downstream sequences on mRNA expression

Although numerous mutations and deletions in the downstream sequence of the gC and other late promoters indicate that this region of the gene is important for maximal expression of a late promoter, the magnitude of the observed effect seems to indicate that a specific downstream sequence modulates but is not absolutely required for late gene expression. To further investigate the influence of the downstream sequence element on promoter activity, synthetic modular late promoters were constructed either with or without the downstream gC sequence (Fig. 7A), and their relative levels of *lacZ* mRNA were determined by primer extension analysis. A promoter containing only the AdML TATA element and the TdT Inr from −5 to +11 (TATA/Inr-11) expressed 76% as much *lacZ* mRNA as the same promoter containing the gC downstream sequence from +20 to +70 (TATA/Inr-3) (Figs. 7A and 7B, lanes 1 and 2), demonstrating that the downstream gC region has a modest effect on expression from a promoter with these particular regulatory elements. The addition of HSV-1 tk downstream sequences from +18 to +53 into TATA/Inr-11 had no discernible effect on the magnitude of *lacZ* expression (Figs. 7A and 7B, lane 3) and did not affect temporal expression from the promoter (data not shown). The data further support the idea that the downstream region of the gC promoter specifically increases the level of mRNA.

Earlier observations have indicated that the UL38 DAS increases expression from promoters containing either weak or strong TATA elements (Guzowski *et al.,* 1994). To investigate whether the gC downstream region might be more influential for mRNA expression from a promoter with a weak rather than a strong Inr element, mRNA expression was measured from modular promoters containing either the AdML TATA element and the dhfr Inr

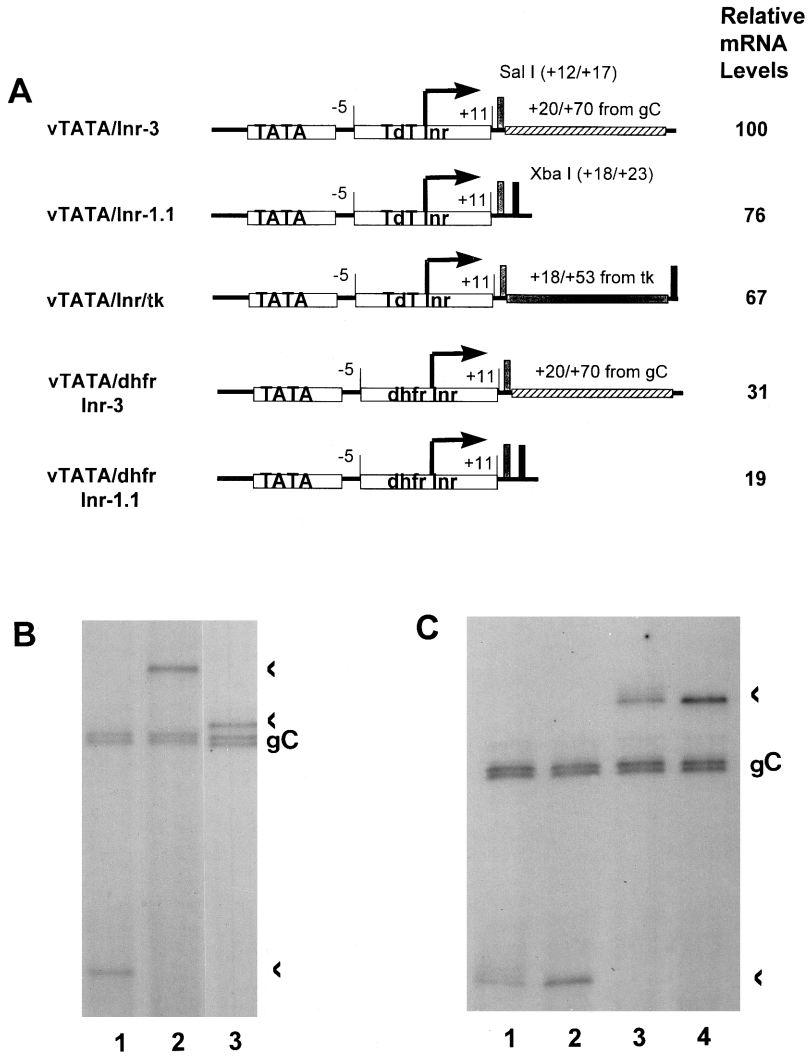


FIG. 7. Addition of the 5′-noncoding region of the gC or tk gene to synthetic promoters containing only TATA and Inr elements. (A) Schematic of the promoters inserted upstream of the *lacZ* gene, and the relative mRNA levels determined by quantification of primer extension assays. (B) Representative primer extension gel showing *lacZ* (arrowheads) and gC mRNA extension products from recombinant viruses vTATA/Inr-11 (lane 1), vTATA/Inr-3 (lane 2), and vTATA/Inr/tk (lane 3). (C) Representative primer extension gel showing *lacZ* (arrowheads) and gC mRNA extension products from recombinant viruses dhfr Inr-11 (lane 1), vTATA/Inr-11 (lane 2), vTATA/dhfrInr-3 (lane 3), and vTATA/Inr-3 (lane 4).

element alone (TATA/dhfr Inr-11) or these two elements in the presence of the gC downstream region from +20 to +70 (TATA/dhfr Inr-3) (Fig. 7A). The *lacZ* mRNA expression was again lower from the promoter without the downstream sequence (Fig. 7C, lanes 1 and 3), but the reduction (39%) was similar to that observed from the promoter containing the TdT Inr (Fig. 7C, lanes 2 and 4). Thus, in the presence of a strong TATA element, the downstream sequence of the gC promoter has a similar effect on expression from promoters containing either strong or weak Inr elements but is not absolutely required for expression.

To determine whether the Inr element was absolutely required for late promoter activity, mRNA expression was measured at 6 and 12 h p.i. from a modular promoter consisting of the AdML TATA element and the random Inr sequence 226, with no further downstream elements (vTATA/226 Inr-11) (Fig. 8A). Although the levels of *lacZ* and gC mRNAs increased from 6 to 12 h in vTATA/Inr-11-infected cells (Fig. 8B, lanes 2 and 4), there was no detectable *lacZ* mRNA in vTATA/226 Inr-11-infected cells at either time point (Fig. 8B, lanes 1 and 3). There was no detectable *lacZ* or gC mRNA in the presence of PAA at either time point (Fig. 8B, lanes 5–8). In contrast, a low level of *lacZ* mRNA was detectable at 12 h p.i. from the modular promoter consisting of the ADML TATA, the 226 Inr sequence, and the downstream gC element (vTATA/Inr-8) (Fig. 8C, lane 4). These results indicate that in the absence of a downstream element, a functional Inr, as well as a TATA element, is necessary for late gene expression. The results also suggest that in the absence of a functional Inr element, a downstream activating element is necessary for late promoter activity.

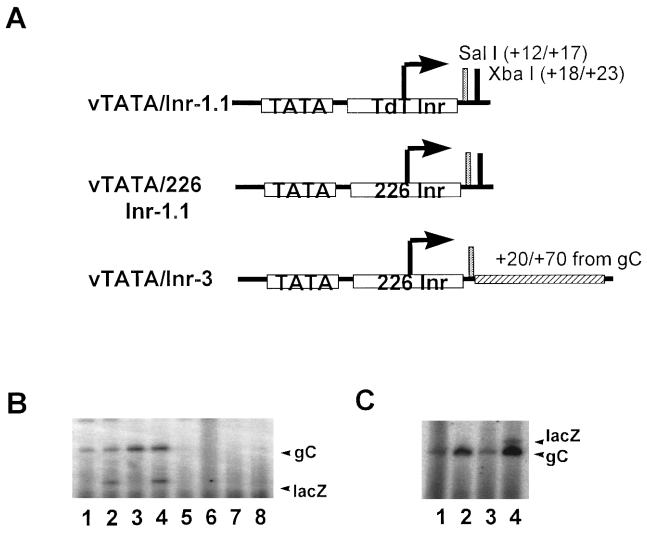


FIG. 8. Expression of mRNA from promoters containing nonfunctional Inr elements. (A) Schematic of the promoters inserted upstream of the *lacZ* gene. (B) Representative primer extension gel showing *lacZ* and gC primer extension products from recombinant viruses vTATA/226 Inr-11 (lanes 1, 3, 5, and 7) and vTATA/Inr-11 (lanes 2, 4, 6, and 8) at 6 h (lanes 1, 2, 5, and 6) and 12 h (lanes 3, 4, 7, and 8) p.i. (Lanes 5–8) From cells infected in the presence of PAA. (C) Representative primer extension gel showing gC and *lacZ* primer extension products from recombinant virus vTATA/226 Inr-11 (lanes 1 and 2) and vTATA/Inr-8 (lanes 3 and 4) at 6 h (lanes 1 and 3) and 12 h (lanes 2 and 4) p.i.

DISCUSSION

In general, promoters for HSV genes resemble other well-characterized eukaryotic promoters for protein-encoding genes (for reviews, see Roizman and Sears, 1996, and Wagner *et al.,* 1995). Although this is not surprising because HSV genes are transcribed by the cellular RNA polymerase II, it is not clear how the eukaryotic transcription machinery differentiates among IE, early, and late genes to regulate the temporal expression of the different classes of viral genes. The virus transactivator VP16 interacts, via specific cellular proteins, with a TAATGARAT *cis*-acting regulatory sequence in the upstream region of IE promoters to activate transcription of these genes. In contrast, although dissection of both early and late virus promoters has identified *cis*-acting regulatory elements necessary for promoter activity, none of these elements has yet been directly associated with viral proteins. Furthermore, the role of these regulatory elements in determining temporal expression is still unclear. Nevertheless, HSV-1 early and late promoters differ in the composition and arrangement of their identified DNA regulatory elements. Early promoters have a TATA element and further upstream regulatory sequences such as Sp1 and CCAAT boxes, whereas late promoters do not appear to have any regulatory elements upstream from the TATA element. Instead, several studies have shown that sequences downstream from the TATA element are required for late gene expression (Guzowski and Wagner, 1993; Huang and Wagner, 1994; Mavromara-Nazos and Roizman, 1989; Weir and Narayanan, 1990).

One regulatory element identified in the region downstream from the TATA element of several late promoters resembles, both in location and sequence homology, the initiator element present in many eukaryotic promoters (Smale and Baltimore, 1989). Indeed, it has been observed that there is no obvious common consensus sequence that distinguishes an HSV late promoter Inr-like sequence from those identified in other eukaryotic systems (Guzowski *et al.,* 1994). In addition, the gC Inr region has been shown to direct transcription *in vitro*, particularly in the presence of ICP4 (Gu and DeLuca, 1994). These observations have indicated that a eukaryotic Inr element might be an integral component of HSV late promoters. To further investigate the functional equivalence of the Inr elements of HSV-1 late promoters and those of other eukaryotic initiators, we constructed model promoters from well-defined regulatory elements of non-HSV genes. These model promoters were inserted into the viral genome and evaluated for their ability to function as late promoters. The results show that certain eukaryotic promoters containing a strong initiator element and a TATA element can function as a late viral promoter. Construction of a several functional late promoters without any HSV sequences reinforces the concept that late promoters do not contain virus specific promoter elements but rather are composed of common eukaryotic transcriptional regulatory elements. However, even the strongest Inr sequence tested, that of the murine TdT promoter, was not sufficient for promoter activity in the context of its natural TATA-less promoter. Similarly, in a previously reported study, insertion of the TdT Inr into a TATA-less tk promoter had no effect on expression (Cook *et al.,* 1995).

The sequence determined for the initiator element has the loose consensus $PyPyA_{+1}NT/APyPy$ (Javahery *et al.,* 1994; Kaufmann and Smale, 1994). The relative strength of late promoters containing different Inr elements roughly corresponded to the previously reported *in vitro* activity of the Inr, with the TdT Inr being the strongest identified initiator element (Javahery *et al.,* 1994; O'Shea-Greenfield and Smale, 1992). The HSV-1 gC initiator element, although not as active as the TdT or AdML Inr elements, is as active as several other known Inr elements. The random sequence designated as 226 has essentially no *in vitro* Inr activity (Javahery *et al.,* 1994) and was not effective as a late promoter element in the viral genome. The sequence at the start of tk transcription also was not able to substitute for a late promoter element. A recent report demonstrated that mutations in the region at the start of tk transcription had little effect on expression of tk mRNA (Cook *et al.,* 1995). Similarly, a hybrid gC promoter containing the tk start region was poorly activated by ICP4 in *in vitro* transcription reactions

compared with the authentic gC promoter (Gu and DeLuca, 1994). These observations suggested that the early tk promoter does not contain a functional initiator element. The inability of this region of the tk sequence to replace other Inr elements in model late promoters is consistent with those suggestions.

Although the initiator element appears to be an integral regulatory element of at least several late HSV-1 promoters as well as many other eukaryotic promoters, the identity of the factors that interact with this element and how such interactions serve to activate transcription are not known. *In vitro* transcription reactions using the gC promoter have indicated that ICP4 activation requires the initiator region, but ICP4 binding at the initiator has not been reported (Gu and DeLuca, 1994). In other systems, several proteins have been reported to bind to the Inr element (Du *et al.,* 1993; Means and Farnham, 1990; Roy *et al.,* 1991; Seto *et al.,* 1991), but the sequence specificity of binding does not appear to correlate exactly with sequences necessary for *in vitro* Inr activity. The loose consensus sequence determined for the Inr suggested that an Inr-binding protein or proteins might bind to this region with relatively low affinity (Javahery *et al.,* 1994). Recent reports suggest that one or more components of the transcription factor IID complex recognize and direct transcriptional regulation through the Inr (Kaufmann *et al.,* 1996; Kaufmann and Smale, 1994; Verrijzer *et al.,* 1994, 1995).

The role of sequences further downstream from the initiator region in late gene expression is even less clear than the role of the initiator. In addition to the present study, several previous studies have shown that deletions and mutations in the nontranslated region of late genes result in reduced gene expression (Guzowski and Wagner, 1993; Homa *et al.,* 1988; Steffy and Weir, 1991; Weir and Narayanan, 1990). In one case, specific mutations in the downstream region of the late UL38 promoter reduced expression by 90% (Guzowski and Wagner, 1993). However, the expression data from recombinant virus vTATA/Inr-11 clearly show that a strong TATA and initiator element can function as a late promoter; the addition of downstream sequences from the gC gene increased expression only moderately. It seems likely that there is a wide variation in the relative strength of individual downstream elements, just as there is for other *cis*-acting regulatory elements. In fact, when gC sequences from +20/+38 were substituted into the UL38 promoter at the downstream DAS element, expression was reduced by >50% (Guzowski *et al.,* 1994).

Inspection of late promoter downstream sequences has yet to reveal a consensus sequence with homology to any known transcription-factor binding sites. However, a cellular protein of ~35 kDa has been identified that binds to the downstream region of the UL38 promoter (Guzowski *et al.,* 1994). The core element of the binding sequence was identified as GGAGCG, and similar sequences were noted in other HSV-1 late genes, including gC, although the gC homology is fairly weak. In addition, this sequence in the gC promoter can be mutated (Steffy and Weir, 1991) or deleted (v23/21 and v27/21) without an apparent negative effect on expression. Further investigation will be required to resolve the identity of factors that recognize this region of the gC gene.

The data from deletion and insertion experiments seem to indicate that the increase in expression due to the downstream gC region, however modest, is sequence specific. Two separate 10-bp deletions in the downstream region, 31/27 and 27/21, resulted in greatly increased levels of mRNA, initiated at the normal start site of transcription. These results further emphasize the potential of the downstream region to affect the extent of mRNA expression. One possible explanation for these observations is that a positive-acting sequence further downstream has been brought closer to the initiator and/or TATA elements, increasing the ease of interaction between the proteins that interact with those elements. That two different deletions, both encompassing one turn of the DNA helix, increased expression is consistent with this possibility, as is the observation that a random 10-bp replacement in one of these deletions lowered expression again. The data obtained with three insertions are also consistent with this possibility. Thus, insertion of a half turn of the helix lowers expression by either moving a positive *cis*-acting sequence further away from the upstream elements or turning the element to the opposite side of the helix. Similarly, insertion of a full turn of the helix also moves the putative downstream element further downstream interfering with its ability to activate expression. Following this line of reasoning, movement of the downstream element two full turns might actually remove stearic constraints and increase the efficiency of interaction again. Of course, until such factors, their binding sites, and their interactions with other factors are identified, this scenario is purely speculative. The possibility also exists that small deletions and insertions have unpredictable effects on the stability of individual transcripts.

The assembly of a promoter from modular units confirms that HSV-1 promoters, particularly late promoters, are fairly simple arrangements of regulatory elements and that these elements are normal eukaryotic regulatory elements. A late promoter may only require two *cis*-acting regulatory elements, a TATA element, and either an Inr element or a downstream activation sequence. In the modular constructions described here, the Inr element appeared to have a much greater effect on expression than the downstream activation sequence. A strong TATA element linked to an Inr element constituted a relatively strong late promoter, and in the absence of a downstream activation sequence, a functional Inr was required for late promoter activity. In the presence of a strong TATA ele-

ment and a functional Inr, the gC downstream regulatory element contributed modestly to the level of mRNA expression but did not influence temporal expression. Interestingly, in the absence of a functional Inr, the addition of the gC downstream element resulted in weak late promoter expression. It remains to be determined whether a late promoter with appreciable strength can be constituted from only a TATA element and a downstream activation sequence. It is likely, however, that there is a great degree of variation in the strength of individual regulatory elements. In a previously reported study, the addition of an upstream regulatory element from an early promoter to a TATA element and a downstream region from a late promoter resulted in a promoter that exhibited properties of both early and late promoters (Mavromara-Nazos and Roizman, 1989). In preliminary studies, we have also found that the addition of upstream elements to a modular late promoter resulted in a promoter that was less sensitive to the effects of PAA (unpublished observations). From such data, it is tempting to conclude that an upstream sequence plus a TATA element equal an early promoter, a TATA and an Inr element equal a late promoter, and all three elements together constitute an early-late (γ_1) promoter. However, in a recent report, the sequence at the start of tk transcription was changed to the TdT initiator sequence. Although higher levels of mRNA resulted, the time course of mRNA expression was unaffected (Cook *et al.,* 1995). Thus, although the overall arrangement of early and late promoter elements may appear simple, the elements themselves, their affinity for various transcription factors, the context of their arrangement, and their interaction with each other may dramatically influence the timing and extent of HSV-1 mRNA expression.

MATERIALS AND METHODS

Cells and viruses

HSV-1 (F) was grown and titered in Vero cells. Recombinant herpesviruses were generated by cotransfection of the appropriate plasmid DNA and HSV-1 (F) genomic DNA into Vero cells using Lipofectamine (GIBCO BRL Life Technologies, Gaithersburg, MD). The virus produced after transfection was plaqued onto Vero cells in the presence of 100 μM acycloguanosine (Sigma Chemical, St. Louis, MO), and multiple individual plaques were selected. Selection of recombinant viruses that expressed the *E. coli lacZ* gene was facilitated by the addition of X-Gal as previously described (Weir and Narayanan, 1988). Individual isolates were purified by at least two more rounds of plaque assay in the absence of acycloguanosine. PCR amplification followed by DNA sequencing was used to verify the promoter region in recombinant viruses.

Construction of plasmids

Routine cloning procedures were similar to those described by Ausubel *et al.* (1987). To construct synthetic promoters containing different initiator elements, the AdML TATA element was synthesized as two complementary oligonucleotides and inserted 5′ to 3′ into the *Sph*I and *Pst*I sites of pGEM3 (Promega, Madison, WI). A second oligonucleotide linker pair containing translational stop codons in all three reading frames was inserted upstream from the TATA oligo into the *Hin*dIII and *Sph*I sites. Different 16-bp initiator elements were inserted downstream from the TATA oligo using the *Pst*I and *Sal*I sites of pGEM3. Each promoter was then isolated as a *Hin*dIII–*Sal*I fragment (5′ to 3′) and inserted into the same sites of pgCP27 (Steffy and Weir, 1991), replacing the gC promoter (−115/+13). The resulting constructions have different Inr elements from −5 to +11, an *Sal*I site from +12 to +17, and sequences from +20 to +70 of the gC nontranslated region beginning at +18 of the new promoter.

The plasmid pTATA/Inr-1 was constructed by insertion of the *Hin*dIII–*Sal*I promoter fragment containing the TdT Inr element into pGal8 insertion plasmid (Weir *et al.,* 1990). To construct pTATA/Inr-11, the 5′ end of the *lacZ* gene in pGal8 was modified to eliminate sequences upstream from the first ATG of the coding sequences. Two complementary oligonucleotides were synthesized to recreate the 5′ end of the *lacZ* gene from the ATG to a *Bam*HI site 42 bp downstream. The 5′ end of the oligonucleotide pair contained an *Xba*I restriction site followed by 2 C residues and then bases corresponding to the *lacZ* coding sequences. This *Xba*I–*Bam*HI fragment was used to replace the *Xba*I–*Bam*HI fragment of pTATA/Inr-1 to generate pTATA/Inr-11, which has an *Sal*I site downstream from the TdT Inr, followed by an *Xba*I site, 2 C bases, and the ATG of the *lacZ* gene. Plasmids pTATA/TdT(−5/+51) and pTATA/AdML(−5/+57) were constructed by substitution of oligonucleotide pairs corresponding to −5/+51 of the TdT promoter and −5/+57 of the AdML promoter, respectively, into the *Pst*I and *Xba*I sites of pTATA/Inr-11. The plasmid pTdT(−50/+51) was constructed by replacement of the *Sph*I–*Pst*I sequence of pTATA/TdT(−5/+51), which contains the AdML TATA element, with an oligonucleotide pair corresponding to sequences from the TdT promoter from −50 to −12.

Plasmids containing deletions in the downstream region of the gC promoter were constructed from the previously described plasmids pgCP21, pgCP23, pgCP26, pgCP27, and pgCP31 (Steffy and Weir, 1991). Each of these plasmids contains the gC promoter from −115 to +70 linked to the *lacZ* coding sequences beginning at +71 relative to the start of transcription. A unique *Hin*dIII site is upstream of −115, and a unique *Bam*HI site is at +42 of the *lacZ* coding sequence. Each of the plasmids has an *Sal*I restriction site (+23/+28, +7/+12, +38/+43,

+14/+19, and +4/+9, respectively) that has been shown to have little effect on *lacZ* expression. Deletions were made by isolating an *Hin*dIII–*Sal*I fragment from one plasmid and linking it to the *Sal*I–*Hin*dIII fragment of another, eliminating bases of the gC promoter between the two *Sal*I sites. Plasmids containing insertions into the downstream gC region were made by PCR amplification of the gC promoter from pgCL5 between −115 and +13 with a random 5 or 10 bp added to the 3′ oligo followed by an *Sal*I restriction site. The *Hin*dIII–*Sal*I fragment then was substituted for the corresponding fragment of pgCP27, which normally has the *Sal*I site at +14/+19. Plasmid 27+23 was made by insertion of a random 23-bp oligonucleotide into the *Sal*I site of pgCP27.

Gene expression assays

To quantify β-gal activity, $\sim 1 \times 10^6$ Vero cells were infected with virus at an m.o.i. of 5 in the presence or absence of 300 μg/ml PAA. At 24 h p.i., the cells were collected by scraping and centrifugation and assayed for β-gal as described (Weir and Dacquel, 1995). The β-gal activity was determined by comparison with standards run in the same assay.

Primer extension assays were done as previously described (Weir and Narayanan, 1988), using either a primer complementary to +15 to +35 of the *lacZ* coding sequences or to +70 to +92 (relative to the start of transcription) of the gC gene. Primer extension products were quantified using a Molecular Dynamics (Sunnyvale, CA) PhosphorImager.

ACKNOWLEDGMENTS

We thank Stephen Smale (University of California, Los Angeles) for plasmid pIG containing the mouse TdT promoter. We also thank Philip Krause and Michael Merchlinsky for helpful discussions.

REFERENCES

Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Smith, J. A., Seidman, J. G., and Struhl, K., Eds. (1987). "Current Protocols in Molecular Biology." Greene and Wiley, New York.

Coen, D. M., Weinheimer, S. P., and McKnight, S. L. (1986). A genetic approach to promoter recognition during trans induction of viral gene expression. *Science* 234, 53–59.

Cook, W. J., Lin, S. M., DeLuca, N. A., and Coen, D. M. (1995). Initiator elements and regulated expression of the herpes simplex virus thymidine kinase gene. *J. Virol.* 69, 7291–7294.

Du, H., Roy, A. L., and Roeder, R. G. (1993). Human transcription factor USF stimulates transcription through the initiator elements of the HIV-1 and the Ad-ML promoters. *EMBO J.* 12, 501–511.

Flanagan, W. M., Papavassiliou, A. G., Rice, M., Hecht, L. B., Silverstein, S., and Wagner, E. K. (1991). Analysis of the herpes simplex virus type 1 promoter controlling the expression of UL38, a true late gene involved in capsid assembly. *J. Virol.* 65, 769–786.

Gu, B., and DeLuca, N. (1994). Requirements for activation of the herpes simplex virus glycoprotein C promoter in vitro by the viral regulatory protein ICP4. *J. Virol.* 68, 7953–7965.

Guzowski, J. F., Singh, J., and Wagner, E. K. (1994). Transcriptional activation of the herpes simplex virus type 1 UL38 promoter conferred by the cis-acting downstream activation sequence is mediated by a cellular transcription factor. *J. Virol.* 68, 7774–7789.

Guzowski, J. F., and Wagner, E. K. (1993). Mutational analysis of the herpes simplex virus type 1 strict late UL38 promoter/leader reveals two regions critical in transcriptional regulation. *J. Virol.* 67, 5098–5108.

Homa, F. L., Glorioso, J. C., and Levine, M. (1988). A specific 15-bp TATA box promoter element is required for expression of a herpes simplex virus type 1 late gene. *Genes and Dev.* 2, 40–53.

Homa, F. L., Otal, T. M., Glorioso, J. C., and Levine, M. (1986). Transcriptional control signals of a herpes simplex virus type 1 late (γ2) gene lie within bases −34 to +124 relative to the 5′ terminus of the mRNA. *Mol. Cell. Biol.* 6, 3652–3666.

Huang, C.-J., and Wagner, E. K. (1994). The herpes simplex virus type 1 major capsid protein (VP5-UL19) promoter contains two cis-acting elements influencing late expression. *J. Virol.* 68, 5738–5747.

Javahery, R., Khachi, A., Lo, K., Zenzie-Gregory, B., and Smale, S. T. (1994). DNA sequence requirements for transcriptional initiator activity in mammalian cells. *Mol. Cell. Biol.* 14, 116–127.

Johnson, P. L., and Everett, R. D. (1986). The control of herpes simplex virus type-1 late gene transcription: a 'TATA-box'/cap site region is sufficient for fully efficient regulated activity. *Nucleic Acids Res.* 14, 8247–8264.

Kaufmann, J., Verrijzer, C. P., Shao, J., and Smale, S. T. (1996). CIF, an essential cofactor for TFIID-dependent initiator function. *Genes Dev.* 10, 873–886.

Kaufmann, J., and Smale, S. T. (1994). Direct recognition of initiator elements by a component of the transcription factor IID complex. *Genes Dev.* 8, 821–829.

Kibler, P. K., Duncan, J., Keith, B. D., Hupel, T., and Smiley, J. R. (1991). Regulation of herpes simplex virus true late gene expression: sequences downstream from the US11 TATA box inhibit expression from an unreplicated template. *J. Virol.* 65, 6749–6760.

Mavromara-Nazos, P., and Roizman, B. (1989). Delineation of regulatory domains of early (β) and late (γ2) genes by construction of chimeric genes expressed in herpes simplex virus 1 genomes. *Proc. Natl. Acad. Sci. USA* 86, 4071–4075.

Means, A. L., and Farnham, P. J. (1990). Transcription initiation from the dihydrofolate reductase promoter is positioned by HIP1 binding at the initiation site. *Mol. Cell. Biol.* 10, 653–661.

Michael, N., and Roizman, B. (1993). Repression of the herpes simplex virus 1 α4 gene by its gene product occurs within the context of the viral genome and is associated with all three identified cognate sites. *Proc. Natl. Acad. Sci. USA* 90, 2286–2290.

O'Hare, P. (1993). The virion transactivator of herpes simplex virus. *Semin. Virol.* 4, 145–155.

O'Shea-Greenfield, A., and Smale, S. T. (1992). Roles of TATA and initiator elements in determining the start site location and direction of RNA polymerase II transcription. *J. Biol. Chem.* 267, 1391–1402.

Panning, B., and Smiley, J. R. (1989). Regulation of cellular genes transduced by herpes simplex virus. *J. Virol.* 63, 1929–1937.

Roberts, M. S., Boundy, A., O'Hare, P., Pizzorno, M. C., Ciufo, D. M., and Hayward, G. S. (1988). Direct correlation between a negative autoregulatory response element at the cap site of the herpes simplex virus type 1 IE175 (α4) promoter and a specific binding site for the IE175 (ICP4) protein. *J. Virol.* 62, 4307–4320.

Roizman, B., and Sears, A. E. (1996). Herpes simplex viruses and their replication. *In* "Fields Virology," 3rd Ed., Vol. 2 (B. N. Fields, D. M. Knipe, and P. M. Howley, Eds.), pp. 2231–2295. Lippincott-Raven, Philadelphia.

Roy, A. L., Meisterernst, M., Pognonec, P., and Roeder, R. G. (1991). Cooperative interaction of an initiator-binding transcription initiation factor and the helix-loop-helix activator USF. *Nature* 354, 245–248.

Seto, E., Shi, Y., and Shenk, T. (1991). YY1 is an initiator sequence-binding protein that directs and activates transcription in vitro. *Nature* 354, 241–245.

Smale, S. T., and Baltimore, D. (1989). The "initiator" as a transcription control element. *Cell* 57, 103–113.

Steffy, K. R., and Weir, J. P. (1991). Mutational analysis of two herpes simplex virus type 1 late promoters. *J. Virol.* 65, 6454–6460.

Verrijzer, C. P., Chen, J.-L., Yokomori, K., and Tjian, R. (1995). Binding of TAFs to core elements directs promoter selectivity by RNA polymerase II. *Cell* 81, 1115–1125.

Verrijzer, C. P., Yokomori, K., Chen, J.-L., and Tjian, R. (1994). Drosophila $TAF_{II}150$: similarity to yeast gene TSM-1 and specific binding to core promoter DNA. *Science* 264, 933–941.

Wagner, E. K., Guzowski, J. F., and Singh, J. (1995). Transcription of the herpes simplex virus genome during productive and latent infection. *Progr. Nucleic Acid Res. Mol. Biol.* 51, 123–165.

Weir, J. P., and Dacquel, E. J. (1995). Plasmid insertion vectors that facilitate construction of herpes simplex virus gene delivery vectors. *Gene* 154, 123–128.

Weir, J. P., and Narayanan, P. R. (1988). The use of β-galactosidase as a marker gene to define the regulatory sequences of the herpes simplex virus type 1 glycoprotein C gene in recombinant herpesviruses. *Nucleic Acids Res.* 16, 10267–10282.

Weir, J. P., and Narayanan, P. R. (1990). Expression of the herpes simplex virus type 1 glycoprotein C gene requires sequences in the 5′ noncoding region of the gene. *J. Virol.* 64, 445–449.

Weir, J. P., Steffy, K. R., and Sethna, M. (1990). An insertion vector for the analysis of gene expression during herpes simplex virus infection. *Gene* 89, 271–274.